How flexible protein structures are? New questions on the protein structure plasticity.

Aurélie Bornot, Bernard Offmann, Alexandre de Brevern

▶ **To cite this version:**

Aurélie Bornot, Bernard Offmann, Alexandre de Brevern. How flexible protein structures are? New questions on the protein structure plasticity.. BioForum Europe, 2007, 11, pp.24. inserm-00189079

HAL Id: inserm-00189079

https://www.hal.inserm.fr/inserm-00189079

Submitted on 11 May 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL author manuscript

How flexible protein structures are?

New questions on the protein structure plasticity.

Aurélie Bornot[1], Bernard Offmann[2,3] & Alexandre G. de Brevern [1]*

[1] INSERM UMR-S 726, Université Paris Diderot, Paris, France
[2] LBGM, Université de la Réunion, Saint-Denis de la Réunion, France
[3] PEACCEL, Technopole, Saint-Denis de la Réunion, France

Contact:
de Brevern A.G., EBGM, INSERM UMR-S 726, Université Paris Diderot, case 7113, 2, place Jussieu, 75251 Paris, France
debrevern@ebgm.jussieu.fr
Tel: (33) 1 44 27 77 31
Fax: (33) 1 43 26 38 30

Aurélie Bornot & Alexandre G. de Brevern

Bernard Offmann

Protein structures and protein structural models are great tools to reach protein function and provide very relevant information for *drug design*. Nevertheless, protein structures are not rigid entities. Cutting-edge bioinformatics methods tend to take into account the flexibility of these macromolecules. We present new approaches used to define protein structure flexibility.

From a rigid to a dynamic view of protein structures

Protein sequences are classically considered as containing the whole information for their three-dimensional (3D) structure. Protein structures are mainly obtained by X-ray or Nuclear Magnetic Resonance (NMR) experiments (see Fig. 1a to 1d) which are listed in the

HAL author manuscript inserm-00189079, version 1

HAL author manuscript inserm-00189079, version 1

Protein Databank (PDB, http://www.rcsb.org/). This experimental information is a critical help for grasping functional properties of proteins, understanding their biological roles, their potential implication in disease mechanisms and for *drug design*. For instance, software as MED-SuMo developed by MEDIT-SA (http://www.medit.fr/), characterize potential binding sites and infer possible biological function of proteins using structural information [**1**].

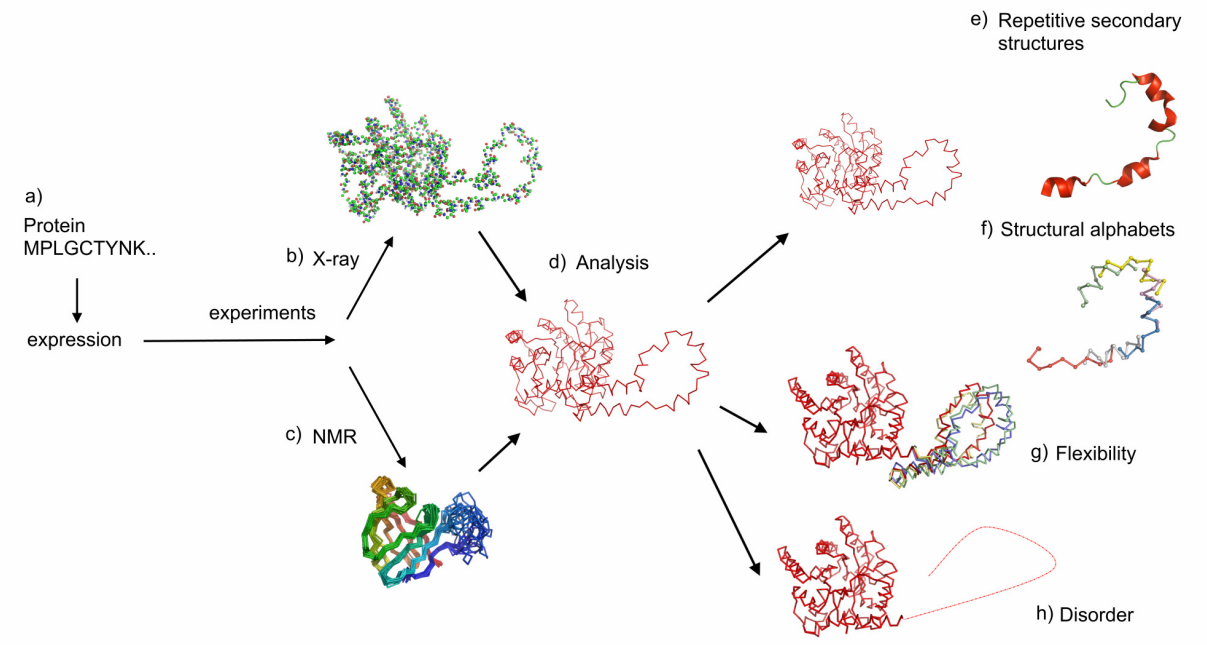

Figure 1. *Different levels of description of the protein structures*.

Costs and difficulties associated to experimental determination of protein 3D structures led to the prediction of relevant structural models from sequence and is becoming a major scientific area. Moreover, the necessity to take into account dynamical properties of proteins to understand their function and mechanism has recently become obvious. Most proteins can exhibit global or local adaptations to perform their function. In the 50s, Karush reported that serum albumin exhibited conformational adaptability for the binding of molecules of very different shapes. Evidence supporting a significant domain movement upon substrate binding was first presented for the association of glucose with hexokinase in 1978.

HAL author manuscript inserm-00189079, version 1

Two experimental methods measure this "flexibility" (see Fig. 1f) in precise regions of protein structures. The atomic mean-square displacement, or *B*-factor, measured during crystallographic experiments reflects an uncertainty about the position of atoms and represents the combined effects of the thermal vibrations and the static disorder. However, *B*-factor is not an absolute quantification of flexibility: it depends on the resolution of the structure, on the refinement procedure, on contacts in the crystal or on the structural environment [**2**]. Flexibility is also indirectly highlighted by NMR experiments that show, in some circumstances, different local conformations that could correspond to different states of protein structures.

The distinction between flexible, highly flexible and unstructured or disordered regions remains a difficult task. A disordered protein, or a disordered region (see Fig. 1g), lacks specific tertiary structure and is composed of an ensemble of conformations, usually with distinct and dynamic ϕ and ψ angles. These regions have low sequence complexity, biased towards polar and charged amino acids and biased away from bulky hydrophobic residues [**3**].

Identification of disordered regions is a hot topic. Some definitions only consider regions that are not stable enough to crystallize. Conversely, others classify loops with high *B*-factors as disordered segments. Similarly, lack of repetitive structures or presence of fluctuating secondary structures are often associated with disordered proteins. Some approaches like heteronuclear multidimensional NMR could provide insight into internal molecular dynamics in the unstructured state [**4**]. Nonetheless, these techniques do not give precise position of the unstructured parts.

New views on this question are now opened: some protein regions can be ordered or disordered depending on the environment and the binding state of the protein, leading to the idea of "dual personality" fragments [**5**].

HAL author manuscript inserm-00189079, version 1

Challenge: Integration of different views of flexibility

In order to describe flexibility in a relevant way to make possible mechanical and functional analyses, we cross results from X-ray and NMR experiments with simulations of protein dynamics using two bioinformatics tools: molecular dynamics and normal modes analysis.

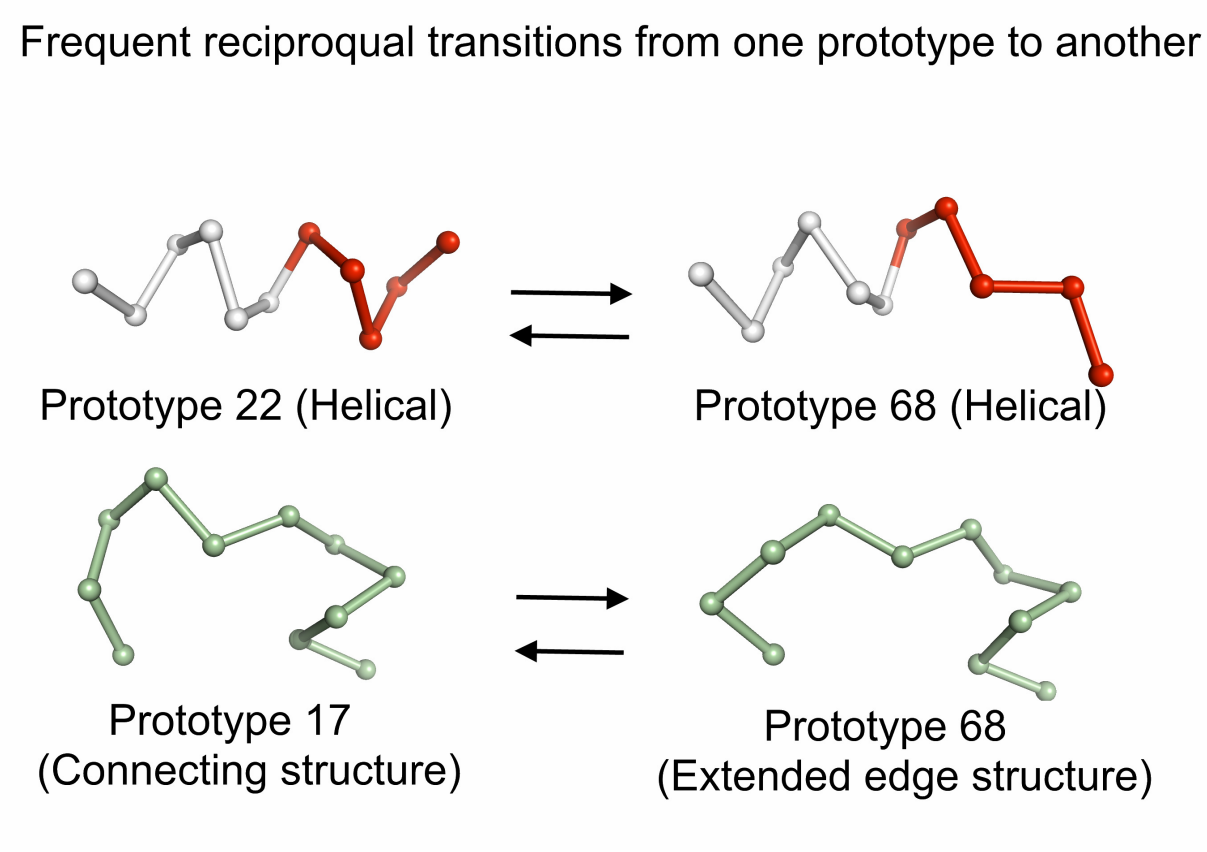

Figure 2. *Examples of local flexibility*.

Analyzing these data requires a fine knowledge and precise description of the available protein folds. Protein folds are often described as a succession of secondary structures (see Fig. 1e). Their repetitive parts (α-helices and β-strands) have been intensively analyzed. Nonetheless, this description does not precisely describe the protein structures, because it fails to describe the relative orientation of connecting regions. Indeed, the coil state covering

HAL author manuscript inserm-00189079, version 1

almost 50% of all residues corresponds to a large set of distinct local protein structures (see Fig. 1f). Thus, to circumvent these difficulties, other approaches were developed. They led to a new approach towards description of 3D protein structures which are now viewed as a combination of small local structures or fragments, also called prototypes [**6**]. For that purpose, we have developed a library of 120 structural prototypes of 11-residue long encompassing all known local protein structures with a good approximation and specially designed for prediction from sequence [**7**]. One of its advantages is to take into account long range interactions thanks to the significant length of the prototypes.

We take advantage of this fine description to observe preferential local deformations of structures (see Fig. 2). In fact, the prototypes describing a given protein structure can be changed during MD simulations or between NMR models, reflecting protein plasticity. Exploration of the conformational space is thus defined as the number of prototypes visited. Successive deformations through time and the propagation of these changes through structures are studied.

Fruitful perspectives: Prediction of flexibility from protein sequence

Using our library, we successfully developed a local structure prediction method from sequence relying on the deduced sequence-structure relationships. We also addressed the question of the structural "predictability" of a sequence and defined indexes aiming at quantifying the sequence informativity on its structure.

We extend now our analysis for deciphering from a sequence the putative flexible zones of a structure and the dynamics of the fragments composing it. For instance, is informativity of the sequence in relation with its structural plasticity? Relying on the integration of multiple views of flexibility, sophisticated analyses are performed for defining new indexes relating sequence to dynamics. These researches would provide new interesting tracks for predicting

flexible regions and putative alternative conformers from sequence.

Acknowledgments

This work was supported by INSERM, Université Paris Diderot and Université de la Réunion. AB benefits from a grant of the Ministère de la Recherche. BO obtained a grant from Conseil Regional de La Réunion and European Union for the PEACCEL project.

Reference

[**1**] Doppelt O., *et. al*., Bioinformation 1, 357-9 (2007).
[**2**] Halle B., Proc Natl Acad Sci 99, 1274-9 (2002).
[**3**] Fink A.L., Cur Op Struct Biol 15, 35-41 (2005).
[**4**] Tompa P., TRENDS in Bioc Sci 27, 527-33 (2002).
[**5**] Zhang Y., *et. al*., Structure 15, 1141-7 (2007).
[**6**] Offmann B., *et. al*., Current Bioinformatics 2, 165-202 (2007).
[**7**] Benros C., *et. al*., Proteins 62, 865-80 (2006).

HAL author manuscript inserm-00189079, version 1